USE OF THE HYPERGEOMETRIC DISTRIBUTION FOR SAMPLING ORDNANCE CONTAMINATED SITES

John N. Lovett, Jr., Ph.D., P.E.
QuantiTech, Inc.
500 Boulevard South, Suite 102, Huntsville, Alabama 35802
Email: jnlovett@quantitech.com
Phone: (256) 650-6263 Fax: (256) 650-5569

Category: Risk Assessment/Site Management

Abstract

In sampling ordnance sites prior to risk assessment using grids, random or structured walks, or other methods of data collection, the question as to how much to sample always arises. The hypergeometric statistical distribution is appropriate in this application, since sampling is from a finite (but unknown) population, specifically the number of unexploded ordnance (UXO) items in a designated sector of the site. The test hypothesis is structured such that an adequate sample from the sector may be chosen to achieve a certain confidence that the actual UXO density for the sector is below an assumed level. In practice, the hypergeometric distribution, though most accurate for this situation, is difficult to calculate (even on the computer) due to potentially large factorial values. QuantiTech has overcome this problem and developed a computer application called "DENSITY ESTIMATOR" which allows the user to input the sector size (in acres or "grids") and varying numbers of UXO items which might be expected in sampling from the sector, beginning with zero. The routine then calculates the probabilities (or confidence levels) that the UXO density in the sector is below a certain level, providing guidelines for initial sampling as well as analysis techniques following data collection. The procedure has proven especially useful in the many instances where no UXO items have been found in sampling from sectors of an ordnance site.

Background

QuantiTech, Inc., has provided sampling planning, support, and/or risk analysis at over 40 ordnance sites throughout the United States. In order to accurately assess risk of exposure to UXO, it is necessary to estimate ordnance density in a designated sector of the site based on a proportionally small (usually less than 10%) sample from the sector. The following two questions arise: 1) How much sampling is "enough?" 2) Once sampling is completed, how may the UXO density in the sector be accurately estimated? QuantiTech has developed a methodology which addresses these questions. The supporting computer implementation is called DENSITY ESTIMATOR.

Methodology

Given a sector or subdivision of an ordnance site, UXO may be scattered randomly throughout the sector or concentrated in certain "hot spots." However, the total (but unknown) number of UXO in the sector is finite and probably small, based on sampling experience. Typically, the sector is divided into grids (although this is not a requirement) and sample grids are chosen randomly for investigation. Magnetic anomalies within each sampled grid are investigated and UXO identified, if present. The results form the basis for a statistical model.

The following symbols may be defined:

N = the number of grids in the sector

n = the number of grids sampled from the sector

d = the number of UXO-contaminated grids in the sector (i.e., the number of sector grids containing at least one UXO) - this is unknown

x = the number of UXO-contaminated grids found in the sample

t = the total number of UXO found throughout all the sampled grids in the sector

The probability of finding x UXO-contaminated grids in the sample, assuming d UXO-contaminated grids in the sector, is given by the hypergeometric probability expression

$$\frac{\binom{d}{x}\binom{N-d}{n-x}}{\binom{N}{n}}, \text{ where a factor of the form}$$

$$\binom{a}{b} = \frac{a!}{(a-b)!b!} = \frac{1\cdot 2\cdot 3\cdots a}{[1\cdot 2\cdot 3\cdots (a-b)][1\cdot 2\cdot 3\cdots b]}$$

Note that as N becomes large, as has been the case in numerous sectors, N! becomes exceedingly large. Indeed, for N = 10, N! = 3,628,800. From field experience and data, it is not uncommon to encounter sectors where N exceeds 10,000. The implementation of DENSITY ESTIMATOR is designed to handle these large numbers with accuracy and speed.

Conclusion

DENSITY ESTIMATOR provides a statistically accurate and convenient means of predicting sector sampling as well as analyzing sample data once collected. It has an advantage in that it can handle a wide range of sector sizes, from the smallest (in some cases less than five acres) to the largest (in excess of 30,000 acres), without loss of accuracy. It may also be used to generate a field sampling table which is simple for field personnel to follow.

References

Cochran, W. G., Sampling Techniques, 2nd Edition. New York: John Wiley and Sons, Inc., 1963.

Duncan, A.J., Quality Control and Industrial Statistics, 4th Edition. Homewood, Illinois: Richard D. Irwin, Inc., 1974.

About The Author

Dr. John N. Lovett, Jr., is a Senior Staff Engineer at QuantiTech, Inc., in Huntsville, Alabama. He received the Ph.D. degree in Industrial Engineering from the University of Arkansas in 1977. Between 1976 and 1984 he was Associate Professor in the College of Engineering at the University of Tennessee at Chattanooga. From 1984 to 1991 he served in the same capacity at the University of Alabama in Huntsville. From 1978 to the present he has provided engineering consultation services to 28 manufacturing, service, and government organizations. Dr. Lovett has taught and/or consulted in areas including statistics, quality assurance and management, operations research, work design, ergonomics, decision theory, and manufacturing processes. He and his wife Janie live in Tennessee where they own and operate an 1873 water-powered mill, museum, and bed-and-breakfast.